CSE574 Introduction to Machine Learning

Support Vector Machine

Jue Guo

University at Buffalo

January 30, 2024

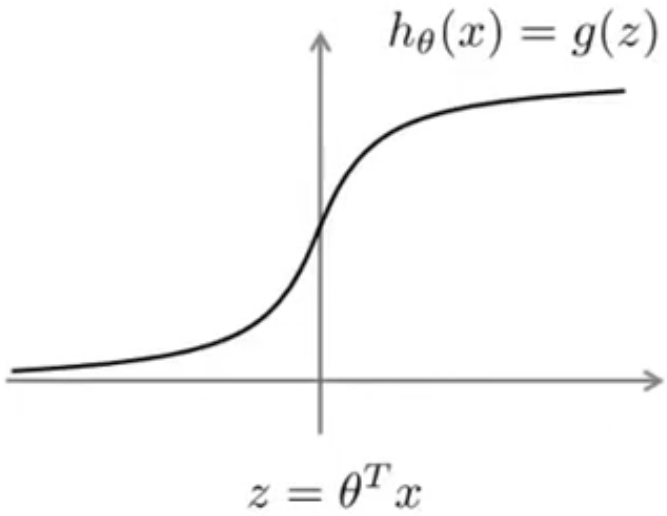
This section is an adaptation of Andrew Ng's Machine Learning Course on SVM.

Outline

Alternative View of Logistic Regression

A quick review: $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

- if $y = 1$, we want $h_\theta(x) \approx 1$, $\theta^T x \gg 0$
- if $y = 0$, we want $h_\theta(x) \approx 0$, $\theta^T x \ll 0$

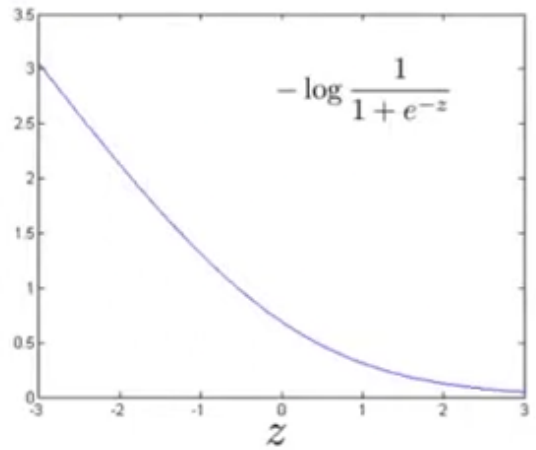


The cost of a single example:

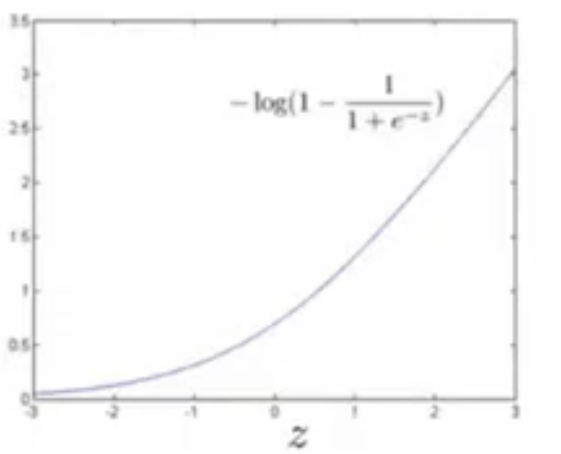
$$
\begin{aligned}
& -\left(y \log h_\theta(x) + (1-y) \log\left(1 - h_\theta(x)\right)\right) \\
= & -y \log \frac{1}{1+e^{-\theta^T x}} - (1-y) \log\left(1 - \frac{1}{1+e^{-\theta^T x}}\right)
\end{aligned}
$$

$$-y \log \frac{1}{1+e^{-\theta^T x}} - (1-y) \log \left(1 - \frac{1}{1+e^{-\theta^T x}}\right)$$

if $y = 1$ (want $\theta^T x \gg 0$)



if $y = 0$ (want $\theta^T x \ll 0$)



Cost Function of Logistic Regression

$$\min_{\theta} \frac{1}{m} \left[\sum_{i=1}^{m} y^{(i)} \left(-\log h_\theta \left(x^{(i)} \right) \right) + \left(1 - y^{(i)} \right) \left(-\log \left(1 - h_\theta \left(x^{(i)} \right) \right) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$$

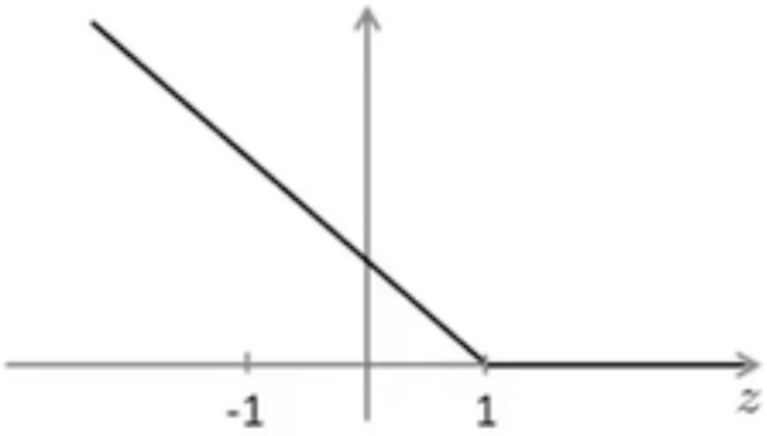
Cost Function of Support Vector Machine

$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$

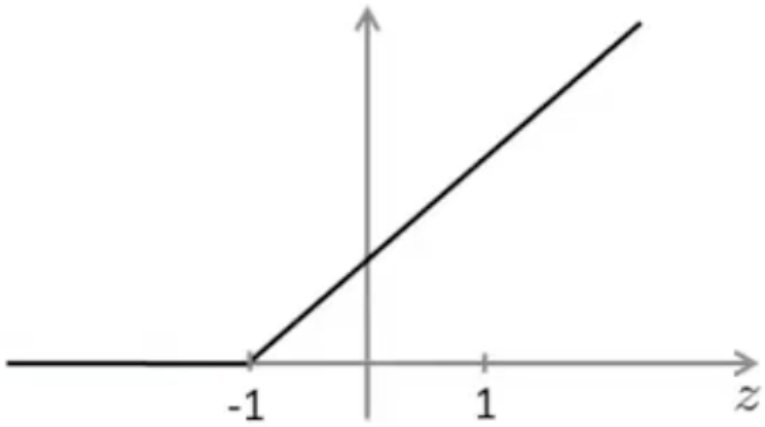
Large Margin Intuition

Support Vector Machine

$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$



If $y = 1$, we want $\theta^T x \geq 1$ (not just ≥ 0)



If $y = 0$, we want $\theta^T x \leq -1$ (not just < 0)

A very large C

Support Vector Machine

$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1 \left(\theta^T x^{(i)} \right) + \left(1 - y^{(i)} \right) \text{cost}_0 \left(\theta^T x^{(i)} \right) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$

Given that C is a very large value, we want that the first term to be 0. Let's try to understand the optimization problem in the context of what would it take to make this first term in the objective equal to 0.

Whenever $y^{(i)} = 1, \theta^\top x^{(i)} \geqslant 1$; Whenever $y^{(i)} = 0, \theta^\top x^{(i)} \leqslant -1$

Now, the optimization problem can be written as:

$$\min C \cdot 0 + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$

$$\text{s.t. } \theta^\top x^{(i)} \geqslant 1 \quad \text{if } y^{(i)} = 1$$

$$\theta^T x^{(i)} \leqslant -1 \quad \text{if } y^{(i)} = 0$$

SVM Decision Boundary: Linearly separable case

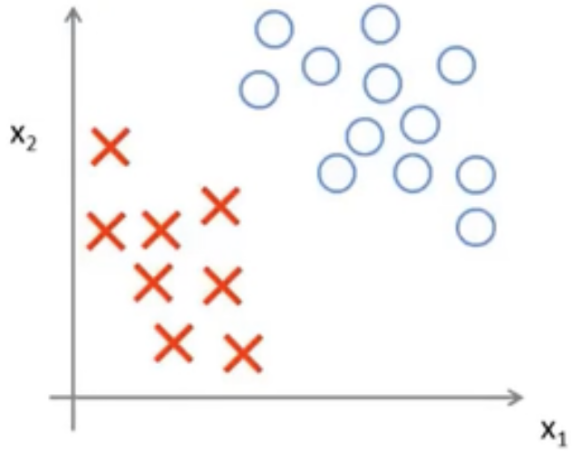
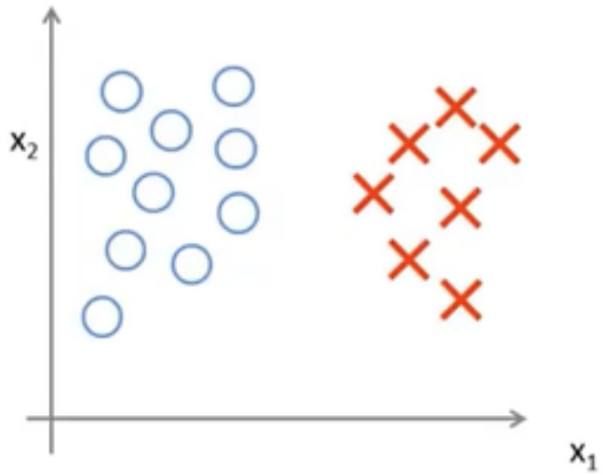


Figure: Linearly Separable Case

Large margin classifier in presence of outliers



Vector Inner Product

Questions?